

Human Skeletal Muscle Nebulin Sequence Encodes a Blueprint for Thin Filament Architecture

SEQUENCE MOTIFS AND AFFINITY PROFILES OF TANDEM REPEATS AND TERMINAL SH3*

(Received for publication, May 15, 1995, and in revised form, November 27, 1995)

Kuan Wang†, Marie Knipfer, Qi-Quan Huang, Ann van Heerden, Lillian Chi-Li Hsu, Gustavo Gutierrez, Xiao-Lan Qian§, and Hansel Stedman§

From the Department of Chemistry and Biochemistry, Biochemical Institute and Cell Research Institute, University of Texas, Austin, Texas 78712 and the §Department of Surgery, Institute for Human Gene Therapy, University of Pennsylvania School of Medicine, Philadelphia, Pennsylvania 19104

Analysis of deduced protein sequence and structural motifs of ~5500 residues of human fetal skeletal muscle nebulin reveals the design principles of this giant multifunctional protein in the sarcomere. The bulk of the sequence is constructed of ~150 tandem copies of ~35-residue modules that can be classified into seven types. The majority of these modules form 20 super-repeats, with each super-repeat containing a 7-module set (one of each type in the same order). These super-repeats are further divided into eight segments: with six segments containing adjacent, highly homologous super-repeats, one single repeat segment consisting of 8 nebulin modules of the same type, and a non-repeat segment terminating with a SH3 domain at the C terminus.

The interactions of actin, tropomyosin, troponin, and calmodulin with nebulin fragments consisting of either repeating modules or the SH3 domain support its role as a giant actin-binding cofilament of the composite thin filament. Such affinity profiles also suggest that nebulin may bind to tropomyosin and troponin to form a composite calcium-linked regulatory complex on the thin filament. The modular construction, super-repeat structure, and segmental organization of nebulin sequence appear to encode thin filament length, periodicity, insertion, and sarcomere proportion in the resting muscle.

Nebulin comprises 2–3% of the myofibrillar protein of skeletal muscle and displays tissue- and development-specific isoforms (600 to 900 kDa) (1–4). It has been demonstrated by immunoelectron microscopy that a single nebulin molecule spans the whole length of the thin filaments with its C terminus anchored at the Z line and its N terminus extending toward the distal ends of thin filaments (2, 3, 5–8). Nebulin has been shown to accompany actin filaments during muscle contraction and extension (2, 5). These data suggest that nebulin forms a composite filament with actin/tropomyosin/troponin, perhaps by binding laterally to actin *in situ*. Significantly, the size of nebulin isoforms is proportional to the length of thin filaments in many skeletal muscles (3, 4). These correlations strongly suggest that nebulin may act as a protein ruler that regulates the length of thin filaments.

Although native nebulin is yet to be isolated and characterized, evidence for nebulin's interactions with actin have been demonstrated by analysis of partial cDNAs encoding nebulin and protein interactions of expressed nebulin fragments. The deduced amino acid sequence of nebulin shows an extensive tandem repeat of ~35-residue modules that are organized into 7-module super-repeats (4, 9, 10, 60). It has been proposed that the 35-residue module is the basic structural unit of the actin binding domains in nebulin and that the super-repeats reflect tropomyosin/troponin binding sites along the nebulin polypeptides (3, 4, 7, 11). Recombinant nebulin fragments containing 2 to 15 modules (7, 12), small native nebulin fragments (13), and 1-module synthetic peptides (11) all bind actin, consistent with this prediction. These experiments indicate that nebulin may contain a string of about 200 actin binding domains along its length. If all sites are operative *in situ*, then nebulin would act as a zipper in its lateral association with actin (12). A one to one matching between nebulin modules with actin protomers would allow nebulin to operate as a protein ruler to determine or stabilize the length of actin filaments (3, 4, 11).

Recent studies on the effect of nebulin fragments on actin-myosin interaction and its regulation by calmodulin raise the intriguing possibility that nebulin might have regulatory functions on active contraction (14). Nebulin fragments bind with high affinity to actin and the myosin head. Fragments from the N-terminal half of nebulin that are situated in the actomyosin overlap region of the sarcomere inhibit actomyosin ATPase activities as well as sliding velocities of actin over myosin during *in vitro* motility assays; while a nebulin fragment near the C terminus, which is localized to the Z line, does not prevent actin sliding. Significantly, calmodulin reverses the inhibition of ATPase and accelerates actin sliding in a calcium-dependent manner. Calmodulin with calcium greatly reduces the binding of nebulin fragments to both actin and myosin. Nebulin may hold the myosin heads close to actin in an orientation that prevents random interaction in resting muscles yet facilitates cross-bridge cycling upon activation by calcium and calmodulin. The data suggest that the nebulin-calmodulin system may function as a calcium-linked regulatory system.

Here, we report the determination and extensive sequence analysis of 5500 amino acids of human fetal skeletal muscle nebulin based on DNA sequencing of five partially overlapping cDNA clones in two open reading frames. Analysis of these sequences, which represent at least 70% of the complete coding sequence, shows that, with the exception of 163 residues at the C terminus, the sequence is arranged as 150 tandem copies of ~35-residue nebulin modules. These modules can be classified into seven types, based on sequence homologies. These modules can be grouped further into 20 super-repeats, with the 7 dis-

* This work is supported by National Institutes of Health Grant AR43514 and grants from the Foundation for Research (to K. W.). The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

† To whom correspondence should be addressed. Tel. and Fax: 512-471-4065; E-mail: kuanw@uts.cc.utexas.edu.

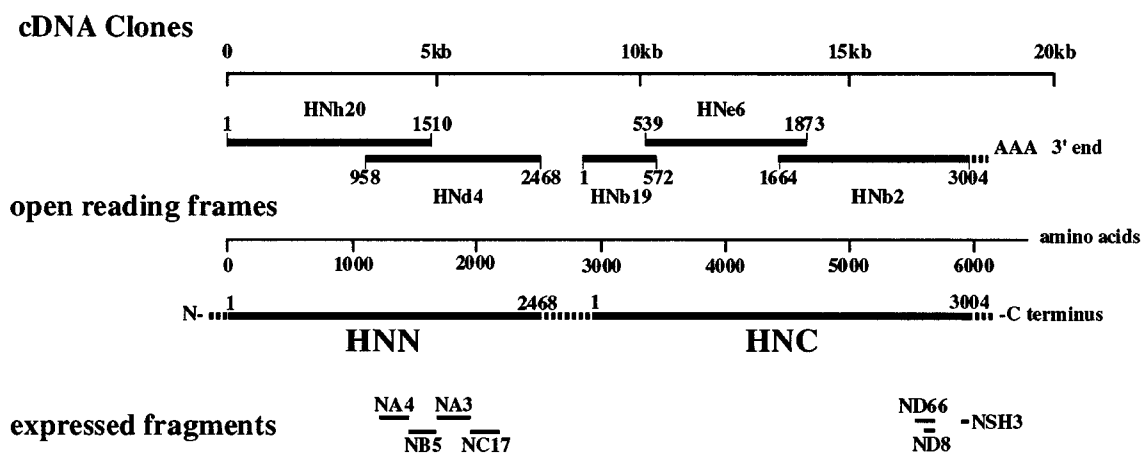


FIG. 1. **Cloning, sequencing, and expression of human fetal nebulin.** Five cDNA clones, HNh20 (4.5 kb), HNd4 (4.5 kb), HNh19 (1.7 kb), HNe6 (4.0 kb), and HNb2 (4.0 kb) are sequenced. Two open reading frames HNN (2468 residues) and HNC (3004 residues) are deduced, with HNN and HNC representing the N- and C-terminal sides of nebulin. Sequence numbers are assigned for each open reading frame, and the corresponding numbers for each cDNA clone are indicated. Seven nebulin fragments, NA4, NB5, NA3, NC17, ND66, ND8, and NSH3 are expressed, with their approximate positions within each open reading frame indicated (refer to Table I for details).

tinct types of modules in each super-repeat, plus a single repeat region containing 8 modules of the same type. Moreover, these super-repeats can be grouped further into six segments, each containing a small number of adjacent, highly homologous super-repeats. The C terminus of nebulin is distinct and contains a Src homology domain 3 (SH3) (15, 16).

The sequence analysis and protein binding studies of expressed nebulin fragments to be presented below suggest that the nebulin sequence encodes the blueprint for the structural and functional compartmentation of thin filaments in the sarcomere of skeletal muscles. The conclusions and implications of nebulin sequence analysis have been presented in a preliminary form (60).

MATERIALS AND METHODS

Isolation of Overlapping cDNA Clones

Human fetal nebulin cDNA clones were isolated from three human fetal skeletal muscle libraries: oligo(dT)-primed cDNA λ gt10 and λ gt11 libraries were used in the original screening (9), and a size-fractionated λ gt10 library of Koenig *et al.* (17) was used in later series of transcript-walking experiments (18).

The initial screening of λ gt11 library led to the identification of two independent λ recombinants λ HN1 and λ HN2 containing nebulin cDNA fragments HN1 and HN2. Both are used as probes to perform "transcript walks" in the λ gt10 libraries. Partial sequence analysis and restriction maps established the order and extent of overlap. Although the two walks have yet to overlap with each other, the outermost cDNA fragments all localize to human chromosome 2 (9), ruling out ligation artifacts in the original cDNA cloning. Five large cDNA clones: HNh20, HNd4 (from HN2 walks) and HNh19, HNe6, and HNb2 (from HN1 walks) were selected for subcloning into pBluescript SK+ (Stratagene) and sequencing (Fig. 1).

DNA Sequencing and Sequence Analysis

Ordered deletions were produced by exonuclease III digestion with a nested deletion kit (U. S. Biochemical Corp.). Clones in pBluescript were sequenced using double-stranded DNA. A Sequenase kit (U. S. Biochemical Corp.) was used for dideoxy sequencing reactions. Each clone was sequenced at least three times in one strand. Sequence analysis was done with either Microgenie (version MG-IM-20, Beckman) or MacVector (version 4.12, Eastman Kodak IBI). Multiple alignment of sequence segments was done with MACAW (version 2.03 Mac 68K, by G. Schuler, NCBI, Bethesda, MD).

Expression and Purification of Cloned Nebulin Fragments

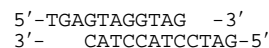
Human nebulin fragments were expressed in *Escherichia coli* as nonfusion proteins by the pET3d expression system as described (19). NA4, NA3, NC17, and ND8 have been described previously and were purified by improvements described in Root and Wang (14). Two addi-

tional fragments near or at the C-terminal region (ND66 and NSH3) were prepared as described below. The molecular parameters of these fragments are summarized in Table I.

ND66—A plasmid bearing a 1.9-kb¹ subclone of the 3'-end clone HNb2 was digested by restriction enzyme *DdeI*, and a 372-bp fragment was purified from 1% agarose gel slice by Prep-a-Gene Matrix (Bio-Rad). This fragment was ligated sequentially to a linear pET3d that was double-digested with *NcoI*-*Bam*HI, first with a synthetic oligonucleotide *NcoI*-*DdeI* initiation adaptor



which directs an in-frame ligation of the 5' end of the coding sequence to the ATG initiation codon in the vector sequence. Then a *DdeI*-*Bam*HI adaptor



containing translation stop codons in all three reading frames was added to join the remaining *DdeI*-cut end of the insert to the *Bam*HI-cut end of the vector. Transformation of BL21 (DE3)pLysS host cells (4 liters) with the resulting plasmid pHNb2D66 led to a high level of expression of soluble ND66 in the cytoplasm upon IPTG induction (0.4 mM IPTG to an $A_{550} = 0.6$ –0.8 culture for 3 h at 37 °C). The bacteria were harvested at 5000 rpm in a Sorvall GSA rotor for 10 min and lysed in a French Press (1,500 p.s.i., 3 times) in 50 ml of lysis buffer (10 mM Na₂P₆, 1 mM EDTA, 1 mM DTT, 2.5 μ g/ml leupeptin, pH 7.0), followed by centrifugation at 14,000 rpm for 20 min in a Sorvall SS-S4 rotor at 4 °C. The supernatant was made 35% saturated in ammonium sulfate at 4 °C for 1 h and spun at 13,000 $\times g$ for 30 min. The pellet was dissolved in 50 ml of lysis buffer and dialyzed overnight, clarified, and applied to a Whatman CM52 column (2 \times 20 cm) equilibrated in lysis buffer. Elution by a linear NaCl gradient (0–1 M NaCl, 150 ml each) yielded 95–99% pure ND66 between 0.35 and 0.40 M NaCl (40 mg per 4-liter culture).

NSH3—A cDNA fragment containing 174 bp of SH3 domain was cloned from a subclone of HNb2 by polymerase chain reaction with primer pairs containing an *NcoI* site and a *Bam*HI site abutting the 5' and 3' ends of the coding sequence, respectively (CTCATCGGATC-CCATGGGAAAAATCTCCGTGCCATG and GTGATGCTTTGGGATC-CCTAAATAGCTTCAACG). The fragment was digested with *NcoI*-*Bam*HI and was ligated to a *NcoI*-*Bam*HI-cut pET3d vector by a one-step reaction with T4 ligase at 4 °C. Transformation of host cells with the recombinant pHNb2SH3 plasmid resulted in the expression of soluble NSH3 upon IPTG induction (0.4 mM IPTG, 3.5 h at 37 °C). NSH3 was purified from a 4-liter culture by lysing host cells in a French Press in 10 mM Tris Cl, 1 mM phenylmethylsulfonyl fluoride, 1 mM EDTA, 0.1 mM DTT, 2.5 μ g/ml leupeptin, pH 7.8, as described above for

¹ The abbreviations used are: kb, kilobase(s); bp, base pair(s); IPTG, isopropyl-1-thio- β -D-galactopyranoside; DTT, dithiothreitol.

TABLE I
Expressed human fetal nebulin fragments

Protein fragment	cDNA clone	Sequence locus	Module	Amino acids	Mass	Isoelectric point
NA4	HNd4	HNN 1209–1450	HNN 34–41	242	27,811	9.24
NB5	HNd4	HNN 1450–1673	HNN 41–47	224	25,866	9.16
NA3	HNd4	HNN 1673–1941	HNN 47–55	269	31,285	9.12
NC17	HNd4	HNN 1941–2184	HNN 55–62	244	28,434	9.07
ND66	HNd4	HNC 2594–2721	HNC 74–78	128	14,403	8.97
ND8	HNb2	HNC 2661–2726 ^a	HNC 76–78	67	7,676	10.08
NSH3	HNb2	HNC 2947–3004	HNC SH3	58	6,437	4.10

^a ND8 contains transposed sequence modules and the expressed fragment consists of ME followed by 2691 to 2726 and 2661 to 2690 (19).

ND66. The supernatant was loaded directly to a DE52 column (2×20 cm) equilibrated in 10 mM Tris Cl, 1 mM EDTA, 0.1 mM DTT, pH 7.8. NSH3 was eluted between 0.4 and 0.6 M NaCl with a 0–1.0 M NaCl gradient (150 ml each). Pooled fractions were dialyzed against 10 mM KP_i, 0.1 mM DTT, pH 7.0 at 4 °C, and applied to a hydroxyapatite column (Bio-Gel HTP, 1×10 cm) at 4 °C. The flow-through fractions containing NSH3 were collected and loaded to an Accel QMA column (1×10 cm) (Millipore) equilibrated in the same buffer and then eluted with an 0–1.0 M NaCl gradient (150 ml each) at 4 °C. NSH3 was eluted between 0.20 and 0.35 M NaCl at 99% purity (40 mg per 4-liter culture).

SDS-Gel Electrophoresis

Proteins were analyzed by either Laemmli gel ($M_r > 20,000$) or a Tris-Tricine gel system (20) ($M_r < 20,000$) and stained with either Coomassie Blue or silver (21).

Quantitative Solid Phase Binding Assays

NA4, NA3, NC17, ND66, ND8, NSH3—Cloned nebulin fragments were dialyzed against a calcium-containing folding buffer (14) either to remove urea and refold (NA4, NA3, NC17) or to exchange buffer (ND66, ND8, NSH3). Prior to coating, fragments were diluted to 5 nM in 10 mM Tris Cl, 150 mM NaCl, pH 7.4 (TBS), added to microtiter plates (Nunc Polysorb FB plates) (100 μ l per well), and incubated overnight at 4 °C. The amount of adsorbed protein is between 7.5 and 15 ng per well, as estimated on a duplicate plate by a copper stain procedure (14). Wells were blocked with 200 μ l of TBS-blocking solution (10 mM Tris Cl, 150 mM NaCl, 0.05% Tween 20, 0.2% bovine serum albumin, pH 7.4) for 1 h at 37 °C. Each adsorbed nebulin fragment was incubated with 100 μ l of actin, tropomyosin, troponin, biotinylated-calmodulin, each ranging from 0.078 μ M to 5.0 μ M, in a binding buffer (10 mM imidazole, 4 mM MgCl₂, 1 mM CaCl₂, 150 mM NaCl, 0.05% bovine serum albumin, pH 7.0) for 2 h at 37 °C. Phalloidin (5 μ M) was present in all actin solutions. After washing, specific antibodies (mouse monoclonal antibodies JLA20 against actin, CH1 against tropomyosin, and JLT12 against troponin (22)) were incubated for 1 h at 37 °C in TBS-blocking solution. Plates were washed three times with TBS-T (10 mM Tris Cl, 150 mM NaCl, 0.05% Tween 20, pH 7.4) and then incubated with a peroxidase-conjugated rabbit anti-mouse antibody (Zymed) for 1 h at 37 °C in TBS-blocking solution, followed by five washes with TBS-T. Color development at 20 °C by 2,2'-azino-bis(3-ethylbenzthiazoline-6-sulfonic acid) and H₂O₂ was monitored at 405 nm every minute with an EIA reader (Bio-Tek model EL-310) and was linear for at least 5 min. Background absorbance from each control well in which binding proteins were deleted is subtracted out prior to plotting and analysis.

To facilitate quantitative analysis, the amount of adsorbed nebulin fragments were kept sufficiently low (7–15 ng/well) so that the total concentrations of the proteins in solution could be approximated as the same as the free protein concentrations. The following hyperbolic equation was derived to fit the data: $A = A_m P / (K_d + P)$; where A is absorbance, A_m is the maximal absorbance at saturation, P is the total protein concentration, and K_d is the microscopic dissociation constant. The stoichiometry of interaction cannot be determined since this method determines only the relative degree, not the absolute amount, of binding.

Biotinylation of Calmodulin—Bovine brain calmodulin (Sigma, 5 mg/ml) in 1 ml of sodium bicarbonate, pH 8.0, was treated while stirring with 100 μ l of biotin-XX, succinimidyl ester (Molecular Probes, B1606) (10 mg/ml in dimethyl formamide) for 1 h at 25 °C. The solutions were dialyzed exhaustively at 4 °C against 10 mM KP_i, 150 mM NaCl, pH 7.2, and then frozen in liquid N₂ and stored at –70 °C.

Binding studies of biotinylated calmodulin to immobilized proteins were done by incubating 100 μ l of biotinylated calmodulin (from 0.030–2.0 μ M) in the binding buffer for 2 h at 37 °C. Plates were washed three

times with TBS-T, followed by incubation with 100 μ l/well streptavidin (Molecular Probes S-888) at 5 μ g/ml in TBS-blocking solution for 15 min at 37 °C. The plates were washed three times with TBS-T and then incubated with 100 μ l/well of biotinylated horseradish peroxidase (Molecular Probes P-917) at 5 μ g/ml in TBS-blocking solution for 30 min at 37 °C. After washing, color development was done as described above.

RESULTS

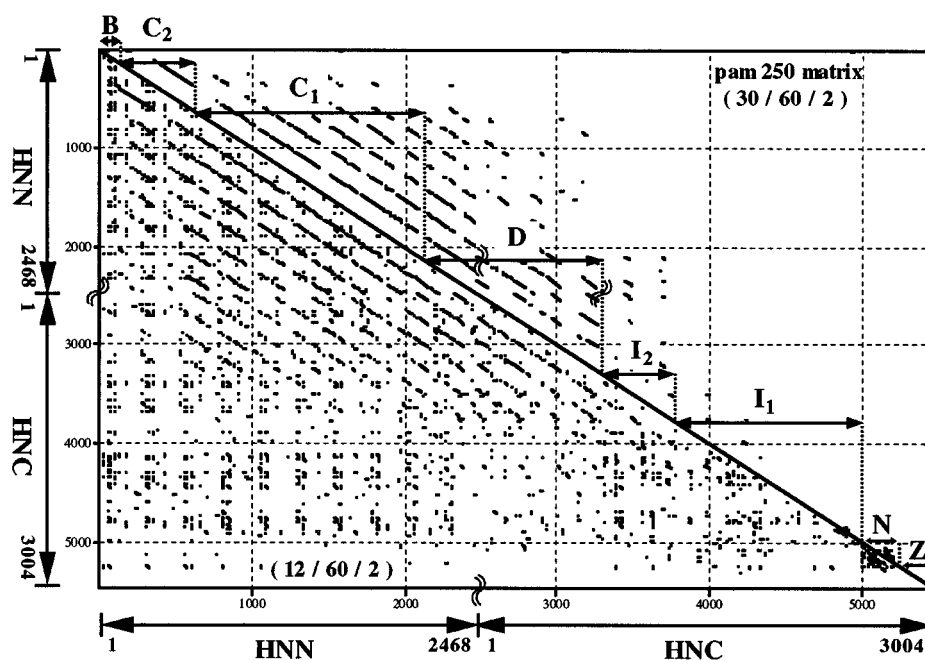
Sequence Motifs of Human Fetal Muscle Nebulin—Partially overlapping cDNA clones were isolated from transcript walking, starting from two independent clones HN2 and HN1 that were obtained from immunological screening of a human fetal muscle λ gt11 library (9). Five clones ranging from 1.7 to 4.5 kb in two groups of overlapping clones (GenBank™ accession numbers U35636 and U35637), were selected for sequence analysis.

A total of ~19 kb has been sequenced to give two open reading frames designated as HNN and HNC with 2468 and 3004 amino acid residues, respectively (Fig. 1). The HNb2 clone contains the 3' end of nebulin transcript including the translation stop codon (TAG), a 422-bp untranslated region and a 42-bp poly(A) tail (data not shown). The 5' walking from HNh20 has yet to give clones that signify the 5' end initiation codon and potential regulatory sequences. The protein sequences encoded by the two open reading frames show extensive tandem repeats of a sequence module that ranges from 31 to 40 residues, with an average of ~35 residues per module. These sequence repeats are easily detected visually by a hexapeptide SXXX(K/R) and a single proline or a smaller cluster of 2–3 prolines that recur every ~35 residues. This module is repeated 69 times in the N-terminal side open reading frame (HNN) and 81 times in the C-terminal open reading frame (HNC), with no obvious linker sequences between modules. We have arbitrarily defined each module as starting at serine of the conserved hexapeptide. This repeating pattern however stops short near the C terminus. The C-terminal segment of 163 residues consists of a linker region of 105 residues enriched in acidic residues, serine, threonines, and a 58-residue SH3 motif which is highly homologous to those found in Src kinases and several cytoskeletal proteins (see below).

Nebulin Modules, Super-repeats, and Segments—Further analysis of sequence homology by protein matrix plots (Pustell protein plot with PAM250 matrix, MacVector) and a multiple alignment program (MACAW with PAM250 matrix) reveals that the majority of these modules can be classified and grouped further into two higher orders of organization: *super-repeats* consisting of seven types of modules and *segments* consisting of contiguous super-repeats that are highly homologous.

The 7-module super-repeat is best visualized by matrix plots with a window size of 8 to 15 with PAM250 scoring matrix. As shown in Fig. 2 (*lower left plot*, with a window size of 12), numerous off-diagonal lines of homologous modules are periodic with a spacing of ~245 residues between successive lines. For a given super-repeat, the intensity of off-diagonal lines

FIG. 2. **Protein matrix plot of human fetal nebulin.** The combined HNN and HNC sequences are scored with PAM250 matrix with two sets of parameters. A 12-residue window (*lower left diagonal*) is used to reveal super-repeats of nebulin modules. A wider, 30-residue window is used to reveal segmental organization of super-repeats (*upper right diagonal*). Segments B, C₂, C₁, D, I₂, I₁, N, and Z are indicated (refer to Table II for details).



along the same vertical axis appears to diminish gradually toward the C terminus. These patterns indicate that nebulin modules are organized into 7-module super-repeats that extend nearly the entire sequences of HNN and HNC. Additionally, the degrees of similarity among nebulin modules are high near the N terminus side and diminish gradually toward the C terminus. The 7 types of modules are designated as *type a* to *type g* in Fig. 3. Each module is designated sequentially from N to C termini as HNN1 to HNC81 (Fig. 3).

Segmental organization of these super-repeats is detected by matrix analysis with a much wider window size that is comparable to the length of nebulin modules. As shown in Fig. 2 (*upper right plot*, with a window size of 30), the off-diagonal lines, 245 residues apart, display staircase steps of various lengths or heights. These patterns indicate a higher degree of homology for super-repeats within each step or segment. Closer examination of local similarity scores by MACAW allows the identification of six segments of homologous super-repeats which are designed as B, C₂, C₁, D, I₂, and I₁ (see "Discussion" for the choice of terminology). Within a given segment, each module type of super-repeats shows a higher similarity score with one another than with those of the same type from other segments. Modules from HNC74 to -81 are all of the same type (*type d*) and, as such, are designated as a single-repeat segment (N segment). Additionally, the C-terminal 163 residues are designated as Z segment, (Table II and Fig. 2).

Nebulin sequence, as presented in Fig. 3, is organized to highlight the 20 seven-module super-repeats that are evident from HNN1 to HNC73. The consensus sequence of each module type, based on a minimum of 50% identity or higher, is annotated on the top of each module and summarized in Fig. 4.

Nebulin Modules and Super-repeats—A consensus sequence of each of the 7 types of modules is deduced by allowing small gaps and identifying conserved residues that appear three or more times at a given position (Fig. 4). Each nebulin module appears to be constructed from two parts: one begins with SXXXY(K/R), that lasts 18–25 residues, followed by a second part starting with a conserved Pro, approximately 13–19 residues long. These conserved residues (Ser, Tyr, and Pro) are useful markers for sequence alignment and are highlighted by reverse contrast in the sequence (Fig. 3). Variations of this

theme occur mostly by substitution or deletion of nonconserved residues. Even for the conserved ones, substitution occasionally occurs for serines and prolines. For example, 12 modules have asparagines substituting for serines in SXXXY. It is noted that Tyr is found at the fifth position in all but 1 module (HNC39 has a SXXY) and is the key landmark for defining the spacing of each nebulin module. The ninth position of each module is also somewhat conserved among the various types. While Tyr is common in modules of *types a* and *b*, aromatic or nonpolar residues such as Trp, Phe (*types f* and *g*), Leu (*types c* and *d*), and Pro (*type e*) are found at this position, characteristic of each type. The sequences following the conserved prolines in the second half are more variable among module types. A shared feature is the presence of one or two K/R at the sixth or seventh position preceding the next SXXXY. The spacings from the conserved Tyr to the conserved Pro of the second half generally fall between 20 and 13 residues for the consensus sequences in Fig. 4: with *type a* modules being the longest at 20 residues; *type g* at 18 residues; *types e* and *f* at 17 residues; *types c* and *d* at 14 residues; and *type b* displaying the shortest at 13 residues. The middle conserved proline is missing in many *type g* modules. In contrast, some of the modules possess two or more prolines in close proximity (*e.g.* *type f* modules and HNC41, -48, and -74–79). This unique distribution of helix-breaking residues may signal distinct folding patterns for these modules. Indeed, a plot of proline distribution along the sequence resembles the tic marks along a ruler with a single proline demarcating each module and multiple prolines (derived from *type f* modules) marking the spacing of each 7-module super-repeat. This unique distribution of prolines has greatly facilitated the identification of super-repeats in the early stage of this work.

Charge Profiles—Another noteworthy feature of nebulin modules is the distribution of the abundant basic and acidic residues. As shown in Fig. 3, in which residues KRH and DE are in *green* and *red*, respectively, most of the conserved Ser, Tyr, and Pro residues adjoin charged groups: Ser is followed by either Asp or Glu. Tyr is flanked by at least one charged group, with KYK (*type a*), EYK (*type b*), XYK (*types c, d, and f*), LYK (*type g*), and KYR (*type e*) being the dominant ones. The ninth position (Y/W/L/P) precedes or follows an acidic group (*types a, b, d, f, and g*) or a basic group (*types c and e*). The conserved Pro

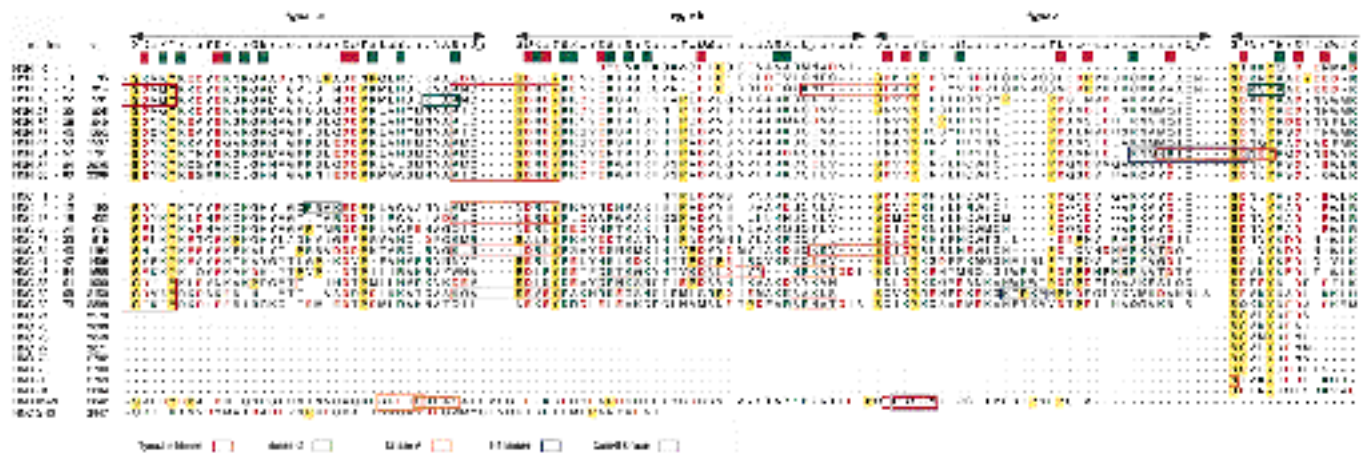


FIG. 3. **Human fetal nebulin sequence.** The sequences of HNN and HNC are grouped into seven types of nebulin modules (*types a to g* on the top of each group) and arranged in 20 super-repeats (from HNN1 to HNC73) plus one single-repeat segment (from HNC74 to HNC81) and C-terminal linker and SH3. The highly conserved SXXXY residues at the beginning and prolines midway through the modules are highlighted with reverse-contrast. Charge residues are in red (DE) and green (HKR), respectively. Subsequences corresponding to protein kinase consensus sites are boxed in various colors. Gaps imposed by sequence alignment are indicated by hyphens (-). The consensus sequence of each type of module is based on 50% identity and is indicated on top of each group. X represents nonconserved residues (<50% identity). + and - represent conserved basic or acidic residues (50% conservation).

is frequently less than two residues away from a D/E. The basic groups generally appear as small clusters of 2 to 4 and alternate with one or two acidic groups. These charge groups are enriched in the SXXXY-containing half of each module. This trend is reversed however for *type d*, where about half of the residues after the conserved Pro are charged ones. Thus, the charge profiles of modules are characteristic of each module type.

The regularity of conserved sequences as well as charge profiles becomes less prominent in HNC, especially for the modules after HNC35. The number and location of prolines and charge group distribution in segments I_2 and I_1 are fairly variable and deviate frequently from the consensus. However, the conserved Tyr and some of the major grouping of basic and acidic groups are still conserved within each type despite sequence variability.

Single Repeats—The regularity reappears in the single repeat segment from HNC74 to -81 near the C terminus. Indeed, the 6 modules from HNC74 to -79 are highly homologous, starting with a unique SSVLY motif and share 12 identical residues (out of 31 per module). These modules are tentatively classified as a *type d* module based mainly on their charge profiles in the second half.

The modules bordering HNC74 and HNC79 are somewhat difficult to classify. HNC70 to -73 are tentatively classified as part of segment I_1 , but appear to be sufficiently homologous to be considered as three single repeats. HNC80 is strikingly similar to a *type d* module HNN1. HNC81 starts with SXXXY, yet without the conserved Pro midway or other conserved residues of *type d* module.

C-terminal SH3 Domain and a Linker Domain—The C-terminal residues (HNC2947–3004) of nebulin shows significant homology with the consensus sequence of SH3 domain that is first identified near the N-terminal noncatalytic region of Src tyrosine kinase. Similar domains have since been found in a variety of enzymes and structural proteins that are important in signal transduction, cortical cytoskeleton, and membrane localization (for reviews, see Refs. 15, 16, and 24). Comparison of nebulin SH3 domain with other SH3-containing cytoskeletal proteins, such as chicken cortactin (p80/p85), human HS1 protein, yeast ABP-1, human amplexin, *Dictyostelium discoideum* myosin 1, and c-Src, indicate extensive homology (Fig. 5). Particularly significant is the presence of highly conserved resi-

dues corresponding to Tyr-90, Tyr-92, Trp-118, Pro-132, and Tyr-135 of c-Src SH3. These hydrophobic residues define the three major hydrophobic binding pockets for proline-rich peptide ligands that bind SH3 (*e.g.* Refs. 25 and 26). Asp-99 of c-Src SH3 is replaced by another acidic residue Glu in this group of SH3s. This conservation of charge is important for SH3-ligand orientation, since the ligand orientation in the binding site is determined by the salt bridge between this acidic residue and arginine of the ligand (26). Another important residue, Tyr-131 in Src SH3, is replaced by Met in nebulin and by Phe, Leu, and Trp in other proteins (Fig. 5). This Tyr residue is situated at one end of the binding site on Src SH3 and is thought to form a fourth pocket that interacts with the flanking residues of the peptide ligand (26). The substitution of a methionine in nebulin suggests that this fourth pocket may be absent or altered to accommodate distinct ligand sequence in the flanking region. On the basis of sequence homology and the observation that the two sequences that correspond to 93–97 and 112–117 of Src SH3 are the same length, it is reasonable to expect that nebulin SH3 would fold similarly into two three-stranded β sheets with two loops from HNC residues 2957 to 2962 and 2976 to 2981 (27).

The 105 residues spanning HNC81 and the SH3 domain is designed as a linker. This linker begins with the SXXXY sequence characteristic of the first half of the nebulin module, but otherwise shares no homology or charge profile with the remaining portion of any of the 7 types of modules. It is highly enriched in acidic residues, serine and threonine totaling 39 mol %.

Segmental Organization and Isoelectric Point Profiles—In addition to similarity scores and charge profiles, another criterion is found useful in identifying the segmental construction of nebulin super-repeats: the profile of isoelectric points of each module along the sequence.

As shown in Fig. 6, the calculated pI values of these modules fall into three classes: basic (8.5 to 10), neutral (6.0 to 7.3), and acidic (4.5 to 5.9). A plot of pI along the sequence revealed striking periodicity throughout most of the HNN region. For example, within HNN modules 3 to 58, a 7-module super-repeat consisting of 5 basic, 1 neutral, and 1 acidic modules is repeated eight times. This pI distribution pattern became less regular from HNN59 to HNC23, with 2–3 basic, 2–3 neutral, and 2–3 acidic modules per super-repeat. From HNC38 to -73,

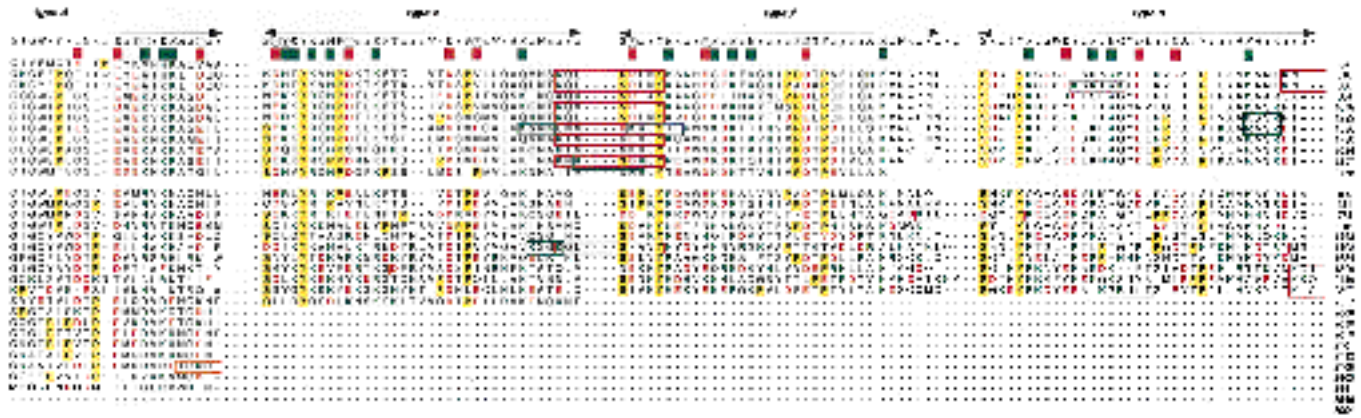


FIG. 3—continued

TABLE II
Segmental organization of human fetal nebulin modules

Segment	Modules	Sequence	Comments
B	HNN 1–2	29–103	Two single modules
C ₂	HNN 3–9	104–351	Two super-repeats
	HNN 10–16	352–598	
C ₁	HNN 17–23	599–842	Six super-repeats
	HNN 24–30	843–1086	
	HNN 31–37	1087–1330	
	HNN 38–44	1331–1574	
	HNN 45–51	1575–1818	
	HNN 52–58	1819–2062	
D	HNN 59–65	2063–2340	
	HNN 66–69 (gap)	2341–2468	Gap may contain additional segments
	HNC 7–13	228–469	
	HNC 14–20	470–712	
	HNC 21–23	713	Only three modules
I ₂	HNC 24–30	813–1057	Two super-repeats
	HNC 31–37	1058–1303	
I ₁	HNC 38–44	1304–1549	Five super-repeats
	HNC 45–51	1550–1794	
	HNC 52–58	1795–2040	
	HNC 59–65	2041–2290	
	HNC 66–72	2291–2542	
	HNC 73	2543–2577	
N	HNC 74–79	2578–2768	Six homologous single repeats at edge of Z line
	HNC 80	2769–2803	Similar to HNN1
	HNC 81	2804–2841	
Z	HNC linker	2842–2951	
	HNC SH3	2952–3004	Z line anchoring domain

most modules are basic, with at most 1 acidic module per super-repeat. The single repeat segment (HNC74 to -81) is mostly basic and neutral with no acidic ones. The linker is nearly neutral ($pI = 6.04$) and the C-terminal SH3 is acidic ($pI = 4.10$). It is striking that this segmental variation in regularity of pI profiles corresponds closely to the staircase-like diagonal matrix plot based on sequence homology (Fig. 2). Taking these independent criteria into consideration, we group these super-repeats into segments that display higher degrees of sequence homology as well as similar pI profiles among the contiguous super-repeats. This segmental organization is illustrated in Fig. 6. Our earlier immunolocalization studies were useful to estimate the distance between HNN and HNC. It is known that HNN41–47 (as an expressed fragment NB5) is localized at $0.88 \mu m$ away from the Z line in adult human quadriceps muscle by site-specific anti-nebulin monoclonal antibody N101 (8). Assuming that fetal and adult nebulin have similar sequences and that each module spans $5.5 nm$ (*i.e.* span of actin subunit), then HNN and HNC is roughly 45 modules apart, corresponding to ~ 1600 residues or $4.7 kb$ of cDNA.

Secondary Structure Propensity—Analysis of nebulin modules for secondary structure propensity led to the prediction of substantial α helical structures. The analysis of a typical super-repeat from HNN16 to -22 by the Chou-Fasman algorithm (MacVector) indicates an average of 32% α helix, 5% β sheet. On the other hand, analysis by the Robson and Garnier algorithm (MacVector) predicts a much higher α helix propensity (70% α helix and 2% β sheet). It is clear that both programs predict high α helix propensity in regions surrounding SXXXXY, especially the preceding sequence (10–15 residues). Indeed, this region of the nebulin module can be induced to form α helix in the presence of anionic detergents and organic solvents (11, 28). In contrast, the regions surrounding the conserved prolines are less regular and devoid of α and β structures.

Phosphorylation Subsequences—The search for protein motifs turns up numerous phosphorylation consensus sites and other protein subsequences. Since nebulin is a major phosphoprotein and undergoes rapid turnover of protein-bound phosphate upon muscle stimulation by cAMP agonists (29), the following potential sites in the sequence are indicated (Fig. 3): tyrosine kinase ((RK)XX(DE)XXX or (RK)XXX(DE)XXY), kinase C ((RK)(ST)X(RK)), cAMP kinase (kinase A) (KRXX(ST) or RRX(ST)), H1 kinase (K(ST)PKK or K(ST)PKK), and Ca^{2+} -calmodulin-dependent kinase II (RXX(ST)(ILVFYW)). The numerous sites for casein kinase 2 and glycogen synthase kinase are not included for the sake of clarity.

It is worth noting that 13 of the 34 tyrosine kinase consensus sites are found in the SXXXXY motif of *type b* modules in HNN and HNC, with 2 to 6 sites in the same region of each of the other types of modules. None is found in the single repeat region. Interestingly, 14 of the 15 cAMP-dependent kinase sites are found in segment I₁, a small adjacent region of segment I₂ (starting at HNC34) and Z segment of HNC, with only 1 in HNN. In the short HNC linker region are localized 3 cAMP kinase sites and 1 Ca^{2+} -calmodulin kinase site, perhaps reflecting its enrichment in serines and threonines. Of the 11 protein kinase C sites, 6 are in *type e* and *f* modules, 0 in *type b* module, and 1 is found in the NSH3 domain at RTGR (HNC residues 2989–2992). It should be noted that actual phosphorylation sites are unknown and may occur at only a small proportion of the possible sites identified in this manner.

Nebulin Homologs—A search of NCBI sequence data bases revealed that homologs of nebulin modules are present in two smaller proteins from human and other species. Cortactin, a cytoskeleton-associated protein substrate (p80/p85) of Src kinase in human platelets, and in chicken and mouse fibroblasts, is composed of $5\frac{1}{2}$ (for a p80 variant) or $6\frac{1}{2}$ (for a p85 variant) tandem repeats of a 37-amino acid module that are linked to a C-terminal SH3 domain by a region rich in proline, serine, and

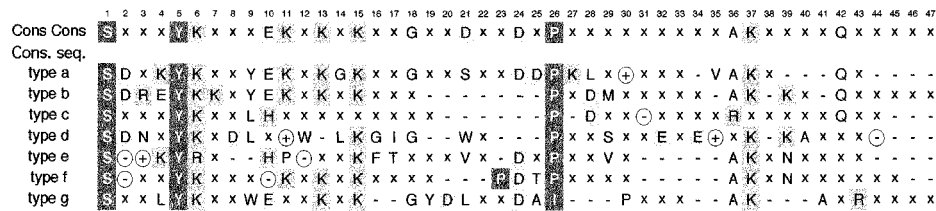


FIG. 4. The consensus sequences of human fetal nebulin modules. The consensus of consensus sequences of seven types of modules (*Cons*) is derived by selecting residues with 40% identity and allowing gaps (-) in order to align conserved SXXXV residues and prolines. X represents nonconserved residues.

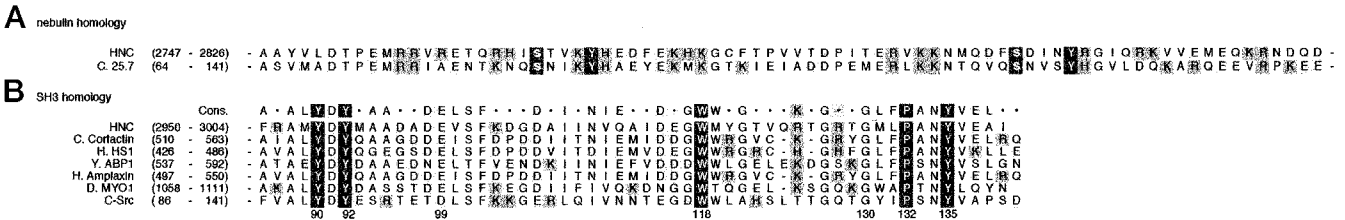


FIG. 5. Nebulin homologs. A, homologs of nebulin modules are found in a hypothetical protein (25.7 kDa) in *C. elegans* (GenBank™ P34417, 221 amino acids). B, SH3 domain homologs. The sequence of nebulin SH3 (HNC) is compared with similar domains from chicken cortactin (GenBank™ G01406, 563 amino acids), human HS1 (GenBank™ P14317, 486 amino acids), yeast ABP1 (GenBank™ A48096, 550 amino acids), *Dictyostelium discoideum* myosin 1D (GenBank™ P34109, 1111 amino acids), and c-Src kinase (GenBank™ P12931, 536 amino acids). Key residues of c-Src SH3 domain are numbered below. The consensus sequence of these SH3s is based on 60% identity and labeled on the top.

threonine (30). All except one of the cortactin modules display the SXXDYK motif characteristics of the *type c* nebulin module, even though the bulk of the sequences is fairly distinct. Interestingly, the tandem repeats bind to F-actin with a stoichiometry of one p80 per 14 actin monomers and a $K_d = 0.43 \pm 0.08 \mu\text{M}$. The SH3 domain of cortactin is highly homologous with that of human nebulin (Fig. 5). Wu and Parsons (30) concluded that SH3 is not directly involved in F-actin binding, based on the lack of cosedimentation of F-actin with a SH3-containing mutant that deletes all tandem repeats. A direct binding study of SH3 with actin, as reported here, however, was not presented.

Two cortactin homologs have been reported. Amplexin, a gene product of *EMS1* gene that is located within the amplified chromosome 11q13 region in human carcinomas (31), exhibits 6½ tandem repeats of cortactin-like modules and an SH3 domain at its C terminus. HS1, a hematopoietic lineage cell-specific protein contains 3½ tandem repeats and a C-terminal SH3 domain (32). None of these repeats, however, contain the SXXDYK motif of cortactin or nebulin modules. These proteins, however, do share the same domain architecture, with tandem repeats of 37-residue modules joined to a C-terminal SH3 by a linker enriched in serine, threonine, and acidic residues (Fig. 7).

Nebulin-like modules are also detected in a 25.7-kDa hypothetical protein encoded by F42H104 of chromosome III of *C. elegans* (GenBank™ P34417) (33). This sequence (c25.7, residue 64–141) is homologous to HNC80 and -81 (HNC residues 2747–2826) with a 40% identity. The possibility exists that this sequence may be part of a larger nebulin-like protein, such as the giant thin filament-associated protein in the body wall muscle of *C. elegans* (34).

Affinity Profiles of Nebulin Fragments toward Myofibrillar Proteins—To identify potential interactions of nebulin with major myofibrillar components, we are performing a systematic screening of protein interactions (affinity profiles) of expressed nebulin fragments by a solid phase binding assay. Seven nebulin fragments in two regions of nebulin sequences were cloned and expressed in *E. coli* as nonfusion proteins (19). These fragments contain 2 to 8 modules each, covering a total of ~1200 residues of nebulin (Table II). The NSH3 domain is

selected based on its sequence homology with SH3s with known three-dimensional structure. Three of these expressed fragments (ND66, ND8, and NSH3) are soluble in the bacterial cytoplasm and are purified by chromatographies in the absence of denaturants. The remaining four (NA4, NB5, NA3, and NC17) are expressed in the inclusion bodies and are solubilized and purified in the presence of urea (19). We have found that the presence of 1 mM Ca^{2+} in the dialysis buffer used to remove urea greatly improves the folding and solubility of these fragments (14).

As a first step, the binding of six nebulin fragments to three major thin filament proteins (actin, tropomyosin, troponin) and calmodulin was studied by solid phase binding assays at physiological ionic strength.

As shown in Fig. 8A, all tested nebulin fragments bind actin. Since phalloidin is included in the buffer to lower the critical concentration of actin polymerization, the binding curves reflect mainly F-actin interaction. The relative affinity of actin binding follows the order: NSH3 ($K_d \sim 0.1 \mu\text{M}$) > ND66, NA3, NA4 ($K_d \sim 0.5 \mu\text{M}$) > NC17, ND8 (nonsaturated at 2 μM actin). This trend is consistent with our previous estimates by cosedimentation studies (NA3, NA4 > ND8) (7, 12, 14). The tight binding of NSH3 to actin is unexpected (see "Discussion") and provides the first evidence that a SH3 domain binds directly to actin.

The binding of tropomyosin to nebulin fragment is generally weaker than actin binding (Fig. 8B). It binds moderately strongly to ND66, NC17, and NSH3 ($K_d \sim 0.5\text{--}1.0 \mu\text{M}$). Its binding to NA3, NA4, and ND8 is not saturated even at 5 μM tropomyosin. Interestingly, the binding of troponin to nebulin fragments is much stronger than tropomyosin-nebulin interaction (Fig. 8C). NA4, NSH3, and ND8 are among the stronger ones ($K_d \sim 0.1 \mu\text{M}$) and NC17, ND66, and NA3 display relative weaker binding ($K_d \sim 0.2 \mu\text{M}$).

Calmodulin (as biotinylated calmodulin), a calcium-mediator of the inhibitory effect of nebulin fragments on actomyosin interaction (14), binds to NA4, NSH3, ND8, and ND66 with higher affinity ($K_d \sim 0.1 \mu\text{M}$) than NC17 and NA3 (unsaturated up to 2 μM calmodulin). This profile thus follows a similar trend as troponin/nebulin affinity, except that the ND66-calmodulin interaction is somewhat stronger. We have previously deter-

FIG. 6. Profile of isoelectric point of nebulin modules and segmental organization of nebulin super-repeats.

The calculated isoelectric points of nebulin modules (as numbered in Fig. 3) are plotted along the HNN and HNC sequences. The gap between HNN and HNC is estimated by immunolocalization of modules HNN41 to -47 with monoclonal anti-nebulin N101 to 0.88 μm from the Z line (8). The segments B, C₂, C₁, D, I₂, and I₁ are presented as boxes of super-repeats. The single repeat region (segment N) and C terminus SH3 and linker (segment Z) are found near the end of HNC. The approximate location of nebulin segments in a rest length (2.3 μm) sarcomere is illustrated on the lower panel. The division of the half-A-band into the M region, P zone, C zone, and D zone is from Sjöström and Squire (35). Two landmarks (NB5 and ND8) were localized by site-specific monoclonal antibodies (N101 and N113) (8). The numbers refer to the individual modules, and the boxes are for the 7-module super-repeats.

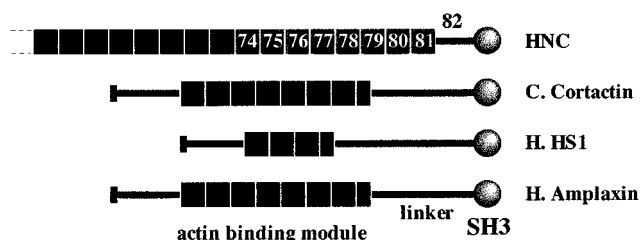
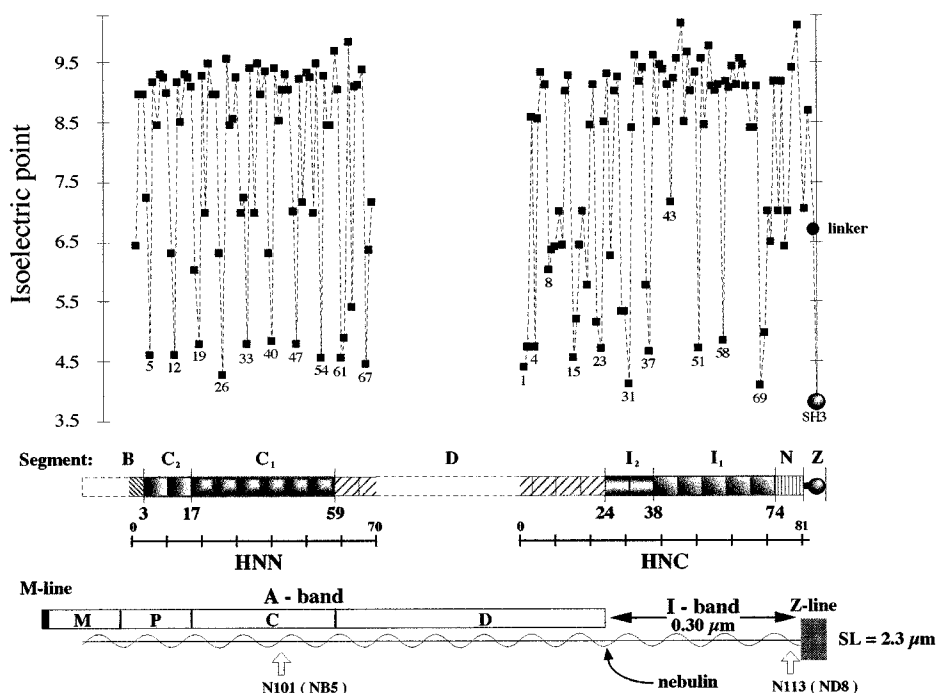


FIG. 7. Domain structure of SH3-containing nebulin homologs. The tandem repeats of 35-residue nebulin modules (HNC) and 37-residue cortactin modules (C. Cortactin, H. HSI, and H. Amphiplexin) are represented by filled squares. The cortactin half-module is represented by a half-square. C-terminal SH3s and linkers are represented as shaded spheres and lines, respectively. Both nebulin and cortactin modules bind F-actin.

mined that NA3 binds to calmodulin with a $K_d \sim 0.6 \mu\text{M}$ in a low ionic strength buffer (14). Additional binding studies have shown that nebulin-protein interactions are fairly sensitive to ionic strength and conditions. Details of these studies will be reported elsewhere.

DISCUSSION

Nebulin as a Blueprint of Thin Filament Architecture and Sarcomere Proportion—Sequence analysis of ~ 5500 residues of human fetal skeletal muscle nebulin reveals a wealth of structural information that can be used to understand its evolution and the design principles of this multifunctional protein in the sarcomere of striated muscles.

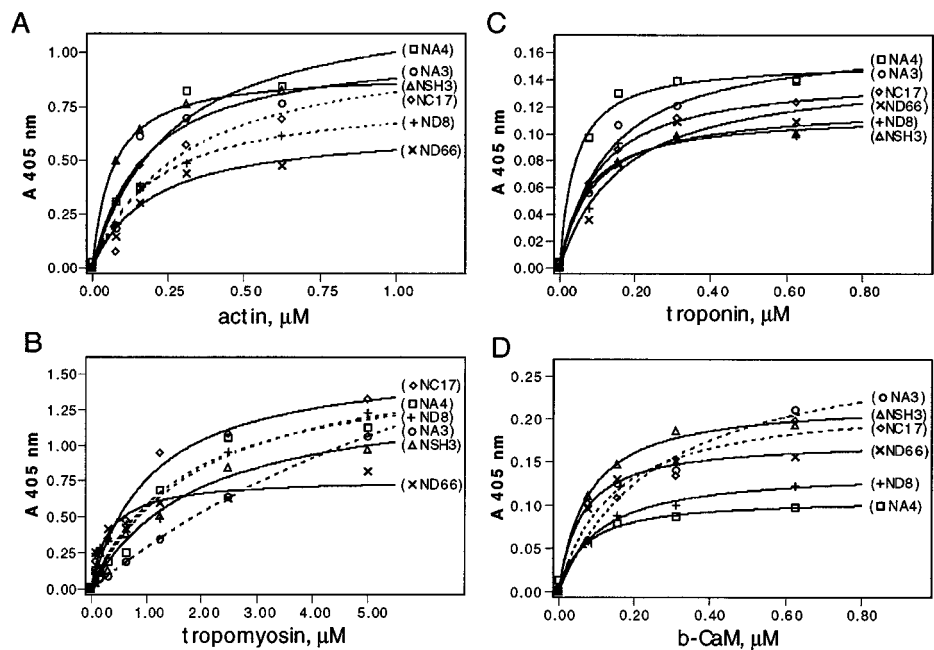
The bulk of the sequence is constructed of more than 150 copies of nebulin modules. Homology analysis of these modules reveal that most of these modules can be classified into seven types and that one of each type forms a 7-module set, to yield 20 super-repeats. Further analysis indicates that the similarity among modules diminishes toward the C terminus. This gradient of diminishing similarity is consistent with the idea that nebulin has evolved from tandem duplication of nebulin modules initiated from its C-terminal end, with the most recently duplicated and conserved ones near the N terminus. We speculate that the total number of duplicated nebulin modules would be determined by the length of thin filaments and the number of actin subunits per helical strand. Further evolution

might have led to the formation of the 7-module super-repeats to provide appropriately spaced sites for tropomyosin and troponin binding, thereby satisfying the spatial constraint of the actin-tropomyosin-troponin complex. The segmentation of super-repeats might be evolved to accommodate its interaction with A-band or I-band components which are themselves segmented in the sarcomere. This speculative idea arose from the following observations. i) The order and span of these segments appear to correlate with the morphological zones within the sarcomere. Close examination of the segmental organization suggests that, provided each super-repeat spans ~ 40 nm, the highly homologous segments of C₁, C₂, and D are expected to overlap with the C zone and D zone of the A-band in a resting muscle sarcomere of 2.3 μm (35). The binding to both actin and myosin by several nebulin fragments in this region (NA3, NA4, ND5, and NC17) supports this notion (14, 19). Interestingly, the combined span of I₁ and I₂ segments near the C terminus would be between 0.30 and 0.50 μm , corresponding to half the width of an I band of a 2.3- to 2.6- μm resting length sarcomere. (ii) The charge profile of modules in HNC display an abrupt transition between the D and I₂ segments (demarcated at HNC24 module, Fig. 5). Most modules in segments I₂, I₁, and N lack the highly regular repeating pattern observed in segments B, C₂, C₁, and D (Fig. 5). This transition is also evident when primary sequences of nebulin modules on either side of this transition are compared (Fig. 3). Modules in the I₂, I₁, and N segments display significant variations, especially in the second half of each module. This is in great contrast to the highly homologous super-repeats in the C₁, C₂, and D segments. We speculate that these sharp transitions between D and I segments might signal the entry of thin filaments into the A band environment of a rest length sarcomere.

Near the C terminus, a short segment of eight tandem repeats of type d modules has been immunolocalized to the edge of the Z line (8) and probably corresponds to the N₁ line to the Z region of the thin filaments which usually appears thicker and stiffer (36). Two nebulin fragments (ND66 and ND8) from this region bind actin but display no affinity toward myosin (14), reflecting a functional distinction from the super-repeat region.

The C-terminal segment of nebulin includes the linker and

FIG. 8. Affinity profiles of nebulin fragments toward thin filament proteins and calmodulin. Nebulin fragments NA3, NA4, NC17, ND66, ND8, and NSH3 (5 nM in 10 mM Tris Cl, 150 mM NaCl, pH 7.4) were adsorbed onto microtiter plates overnight at 4 °C, washed, blocked, and incubated with actin (A), tropomyosin (B), troponin (C), and biotinylated calmodulin (*b*-CaM, D) in a binding buffer (10 mM imidazole, 4 mM MgCl₂, 1 mM CaCl₂, 150 mM NaCl, 0.05% bovine serum albumin, pH 7.0) for 2 h at 37 °C. Phalloidin (5 μ M) was present in actin solutions. The bound proteins were detected with peroxidase-conjugated antibodies (A, B, and C) or streptavidin (D). All solid curves were calculated using a one-class binding model. Dotted lines indicate binding isotherms that have not reached saturation.



an SH3 domain, both of which are distinct from nebulin modules. Significantly, the SH3 domain also binds actin (Fig. 8). This interaction may contribute to the anchoring of nebulin to the Z line (8).

If the basic premises of the foregoing speculative analysis are correct, then the nebulin sequence encodes not only a blueprint for the length and architecture of thin filaments, but also instructions for the degree of overlap of thin and thick filaments and sarcomere length in the resting muscle. Nebulin thus may impart a functional and structural compartmentation along the otherwise uniform actin/tropomyosin/troponin filaments. Remarkably, this task is accomplished mostly by modifying and duplicating a short yet versatile 35-residue building block.

Interaction of SH3 Domain with Thin Filament Proteins—The receptor proteins for SH3 domains in signal transduction pathways are being actively pursued by many laboratories, and recent activities have focused on a series of proline-rich peptides and proteins containing these consensus sequences (15, 24, 25, 37, 38). The receptors for SH3 domains in cytoskeletal proteins, however, have remained obscure. Our demonstration that expressed NSH3 domain of nebulin binds to actin with high affinity provides the first biochemical evidence that SH3 has the potential of binding directly to this key component of cytoskeletal filaments. The interaction perhaps occurs through a site other than the primary one for proline-rich peptides, since these proline-rich consensus peptides are absent in actin. Its interaction with actin suggests that nebulin may terminate at the Z line by binding SH3, and possibly the linker, to either the side or the end of the actin filaments that traverse the entire width of the Z line. It is not yet clear whether NSH3 binds to other Z line proteins such as α actinin (39), Cap Z (40), tensin (41), and titin (42). The binding to tropomyosin and troponin is curious, since both proteins generally have been assumed to be distributed along the length of actin thin filament outside the Z line, with troponin starting at the first thin filament repeat about 80–100 nm away from the edge of the Z line (43–45). The significance, if any, of NSH3-tropomyosin/troponin interactions therefore might lie in its potential regulatory role during myofibrillar assembly in developing muscles.

Whether the N terminus of nebulin also possesses a unique segment that binds the capping protein for the pointed end of

actin filaments such as tropomodulin (46) is yet unclear. The similarity of module HNN1 with the C-terminal end module HNC80 hints that the 5'-most sequence may be nearby the N terminus of HNN.

Super-repeats and Tropomyosin/Troponin Binding—We have recently proposed that nebulin is multifunctional and serves at least a dual role. First, it acts as a structural template or protein ruler to regulate the length of actin filaments (2–4, 8, 12). Second, it serves as a regulatory protein that tethers myosin heads to actin in rigor conditions and releases and accelerates myosin sliding on actin in activating conditions in the presence of calcium/calmodulin and ATP (14). The thin filaments of skeletal muscle thus might be dually regulated by tropomyosin/troponin and by nebulin/calmodulin. Further understanding of the interplay between these two systems would require the knowledge of their relative structural disposition and potential reorganization in rigor, activating, and relaxing states. The observed interaction of nebulin fragments with tropomyosin and troponin, if manifesting *in situ* for most nebulin super-repeats, would suggest an intriguing possibility that nebulin and tropomyosin/troponin may interact to form "composite" regulatory complexes, as depicted in Fig. 9. In this hypothetical model, each 7-module super-repeat of nebulin polypeptides (either monomer or dimer) binds seven actin protomers and one tropomyosin and troponin complex in a parallel fashion to span the width of one thin filament periodicity of ~ 40 nm. Thus, the nebulin super-repeat serves as a signal for the binding, orientation, and spacing of the actin/tropomyosin-troponin complex. It is worth noting that several anti-nebulin monoclonal antibodies label nebulin epitopes that coincide with the ~ 40 nm thin filament periodicity (8, 47). These periodic nebulin epitopes may be located in one of the seven types of nebulin modules. Additional support for this idea of composite regulatory complex is found in our observation that the N terminus of actin can be cross-linked to ND8 by a zero length cross-linker.² This front end of actin subdomain 1 where the N terminus is located has also been implicated as a tropomyosin binding site in the "off" state when tropomyosin blocks myosin head attachment (48, 49). Nebulin, tropomyosin, and troponin

² C. L. Shih and K. Wang, manuscript in preparation.

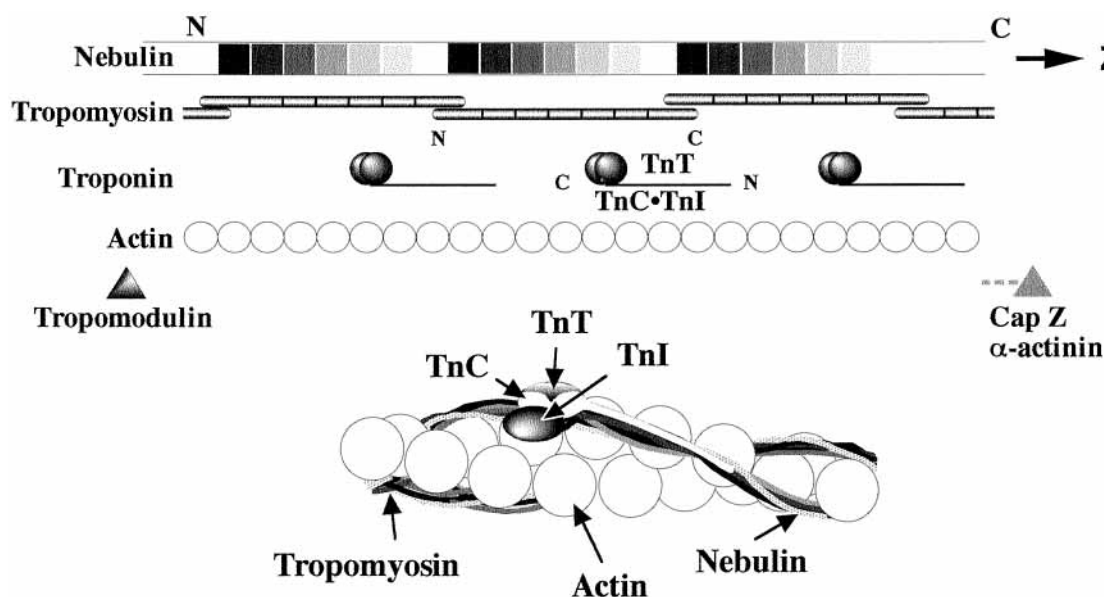


FIG. 9. A hypothetical model of a composite regulatory complex containing nebulin and tropomyosin/troponin on the thin filaments of skeletal muscle sarcomere. In this working model, each seven-module nebulin super-repeat (squares with graded shading) binds one tropomyosin, possibly through the seven charge clusters along the length of each tropomyosin (partitioned rod), and one troponin complex (shaded spheres with a tail), consisting of TnT, TnI, and TnC in orientations as specified by the N and C termini (49). Each nebulin super-repeat binds to seven actin monomers (open circles) along the thin filament. Two strands of this composite regulatory complex are depicted as winding around the outer actin domains (subdomains 1 and 3) along the double helical polymer. Tropomodulin is thought to bind to the pointed end of actin filaments and/or the N terminus of nebulin. Cap Z, α -actinin may bind to the barbed end of actin filaments and/or the C terminus of nebulin.

are probably in close proximity with each other on actin, at least in the off state.

Functional and Structural Analogs of Nebulin—What has emerged from this and other studies is a picture of nebulin that is analogous functionally to caldesmon of smooth muscles (14). Both are putative thin filament-associated regulatory proteins that bind actin, myosin, calmodulin, and tropomyosin and are thought to tether myosin heads to the actin filaments in relaxed muscles and to facilitate myosin-actin interaction in activated muscles (see, for example, Refs. 50–53). Both are phosphorylated *in vivo*. The abundant phosphorylation consensus subsequences in nebulin (see Fig. 3) and the rapid turnover of protein phosphate *in vivo* (29) suggest that, as is the case with caldesmon (54–57), nebulin phosphorylation may be an additional mechanism for modulating the inhibitory effect of nebulin on actomyosin interaction. Differences in molecular properties (size, sequence, and domain organization) may reflect the functional diversity of the thin filament-based regulatory mechanism in smooth muscles (which fine tune contraction) and skeletal muscles (which aim for force and speed).

In non-muscle cells, cortactin and its homologs share the same modular architecture as the C-terminal end of human nebulin (see Fig. 5) and may be considered as functional or structural analogs of a family of SH3-containing proteins which bind actin polymers to SH3 target sites in the cortical cytoskeleton or in the Z line. It would be of great interest to know if the short cortactin tandem repeats regulate the length of the attached actin oligomers or polymers.

Other proteins that contain nebulin modules are also being detected. The presence of nebulin-like modules in a hypothetical protein (25.7 kDa) in *C. elegans* demonstrates that nebulin modules are ancient and conserved. The similarity between these modules with HNC81, found near the C-terminal of human nebulin also point to the importance of nebulin modules in this anchor region. In this connection, it is significant that at least 11 nebulin modules that are highly homologous to (HNC69 to -79) human nebulin are present in cardiac nebulin, a 107-kDa nebulin-like protein found in cardiomyocytes

of human and chicken (58). This anchor region structure may be a theme upon which structural or functional variants evolve in non-muscle and cardiac muscle cells.

Human Adult Nebulin Sequence—The complete sequence of adult human nebulin was described (59) while this manuscript was in review. The adult nebulin sequence (EMBL accession number X83957) consists of 185 modules, with the central 154 copies grouped into 22 super-repeats and unique sequences at both the N and C termini. A detailed comparison of human adult and fetal nebulin sequences will be presented elsewhere.

Acknowledgments—We thank Drs. Russell F. Doolittle, Lawrence L. Poulsen, and James Sylvester for advice on sequence analysis and cloning techniques. We are grateful to John Wright for his expert help in aligning the sequence and preparing the illustrations, and to Qian Ming and Mark Smyer for the expression and preparation of nebulin fragments. We thank Dr. C. Moncman for her input in preparing this manuscript.

REFERENCES

- Locker, R. H., and Wild, D. J. C. (1986) *J. Biochem. (Tokyo)* **99**, 1473–1484
- Wang, K., and Wright, J. (1988) *J. Cell Biol.* **107**, 2199–2212
- Kruger, M., Wright, J., and Wang, K. (1991) *J. Cell Biol.* **115**, 97–107
- Labeit, S., Gibson, T., Lakey, A., Leonard, K., Zeviani, M., Knight, P., Wardale, J., and Trinick, J. (1991) *FEBS Lett.* **282**, 313–316
- Maruyama, K., Matsuno, A., Higuchi, H., Shimaoka, S., Kimura, S., and Shimizu, T. (1989) *J. Muscle Res. Cell Motil.* **10**, 350–359
- Pierobon-Bormioli, S., Betto, R., and Salviati, G. (1989) *J. Muscle Res. Cell Motil.* **10**, 446–456
- Jin, J.-P., and Wang, K. (1991) *FEBS Lett.* **281**, 93–96
- Wright, J., Huang, Q.-Q., and Wang, K. (1993) *J. Muscle Res. Cell Motil.* **14**, 476–483
- Stedman, H., Browning, K., Oliver, N., Oronzi-Scott, M., Fischbeck, K., Sarkar, S., Sylvester, J., Schmickel, R., and Wang, K. (1988) *Genomics* **2**, 1–7
- Zeviani, M., Darras, B. T., Rizzuto, R., Salviati, G., Betto, R., Bonilla, E., Miranda, A. F., Du, J., Samitt, C., Dickson, G., Walsh, F. S., Dimauro, S., Francke, U., and Schon, E. A. (1988) *Genomics* **2**, 249–256
- Pfuhl, M., Winder, S. J., and Pastore, A. (1994) *EMBO J.* **13**, 1782–1789
- Chen, M.-J. G., Shih, C.-L., and Wang, K. (1993) *J. Biol. Chem.* **268**, 20327–20334
- Tatsumi, R., Hattori, A., and Takahashi, K. (1992) *J. Biochem. (Tokyo)* **112**, 780–785
- Root, D. D., and Wang, K. (1994) *Biochemistry* **33**, 12581–12591
- Mayer, B. J., and Baltimore, D. (1993) *Trends Cell Biol.* **3**, 8–13
- Musacchio, A., Gibson, T., Lehto, V.-P., and Saraste, M. (1992) *FEBS Lett.* **307**, 55–61

17. Koenig, M., Monaco, A. P., and Kunkel, L. M. (1988) *Cell* **53**, 219–228
18. Stedman, H., Browning, K., Qian, X.-L., Oliver, N., Oronzi-Scott, M., Fischbeck, K., Sarkar, S., Sylvester, J., Schmickel, R., and Wang, K. (1989) in *Cellular and Molecular Biology of Muscle Development* (Kedes, L., and Stockdale, F. E., eds) pp. 181–196, Alan R. Liss, Inc., New York
19. Jin, J.-P., and Wang, K. (1991) *J. Biol. Chem.* **266**, 21215–21223
20. Schägger, H., and von Jagow, G. (1987) *Anal. Biochem.* **166**, 368–379
21. Gottlieb, M., and Chavoko, M. (1987) *Anal. Biochem.* **165**, 33–37
22. Lin, J. J.-C. (1981) *Proc. Natl. Acad. Sci. U. S. A.* **78**, 2335–2339
23. Wang, K., McCarter, R., Wright, J., Beverly, J., and Ramirez-Mitchell, R. (1993) *Biophys. J.* **64**, 1161–1177
24. Bar-Sagi, D., Rotin, D., Batzer, A., Mandiyan, V., and Schlessinger, J. (1993) *Cell* **74**, 83–91
25. Ren, R., Mayer, B. J., Cicchetti, P., and Baltimore, D. (1993) *Science* **259**, 1157–1161
26. Feng, S., Chen, J. K., Yu, H., Simon, J. A., and Schreiber, S. L. (1994) *Science* **266**, 1241–1247
27. Musacchio, A., Noble, M., Pauptit, R., Wierenga, R., and Saraste, M. (1992) *Nature* **359**, 851–855
28. Chen, J. K., and Schreiber, S. L. (1994) *BOMC (Bull. Osaka Med. Coll.) Lett.* **4**, 1755–1760
29. Somerville, L., and Wang, K. (1988) *Arch. Biochem. Biophys.* **262**, 118–129
30. Wu, H., and Parsons, J. T. (1993) *J. Cell Biol.* **120**, 1417–1426
31. Schuurin, E., Verhoeven, E., Mooi, W. J., and Michalides, R. J. A. M. (1992) *Oncogene* **7**, 355–361
32. Kitamura, D., Kaneko, H., Miyagoe, Y., Ariyasu, T., and Watanabe, T. (1989) *Nucleic Acid Res.* **17**, 9367–9379
33. Wilson, R., Ainscough, R., Anerson, K., Baynes, C., Berks, M., Bonfield, J., Burton, J., Connell, M., Copsey, T., Cooper, J., Coulson, A., Craxton, M., Dear, S., Du, Z., Durbin, R., Favello, A., Fraser, A., Fulton, L., Gardner, A., Green, P., Hawkins, T., Hiller, L., Jier, M., Johnston, L., Jones, M., Kershaw, J., Kirsten, J., Laisster, N., Latreille, P., Lightning, J., Lloyd, C., Mortimore, B., O'Callaghan, M., Parsons, J., Percy, C., Rifken, L., Roopra, A., Saunders, D., Showkeen, R., Sims, M., Smaldon, N., Smith, A., Smith, M., Sonnhammer, E., Staden, R., Sulston, J., Thierry-Mieg, J., Thomas, K., Vaudin, M., Vaughan, K., Waterston, R., Watson, A., Weinstock, L., Wilkinson-Sproat, J., and Wohldman, P. (1994) *Nature* **368**, 32–38
34. Goetinck, S., and Waterston, R. H. (1994) *J. Cell Biol.* **127**, 79–93
35. Sjöström, M., and Squire, J. M. (1977) *J. Mol. Biol.* **109**, 49–68
36. Franzini-Armstrong, C. (1970) *Tissue & Cell* **2**, 327–338
37. Yu, H., Rosen, M. K., Shin, T. B., Seidel-Dugan, C., Brugge, J. S., and Schreiber, S. L. (1992) *Science* **258**, 1665–1668
38. Yu, H., Chen, J. K., Feng, S., Dalgarno, D. C., Brauer, A. W., and Schreiber, S. L. (1994) *Cell* **76**, 933–945
39. Nave, R., Furst, D. O., and Weber, K. (1990) *FEBS Lett.* **269**, 163–166
40. Casella, J. F., Craig, S. W., Maack, D. J., and Brown, A. E. (1987) *J. Cell Biol.* **105**, 371–379
41. Lo, S. H., Janmey, P. A., Hartwig, J. J., and Bo, C. L. (1994) *J. Cell Biol.* **125**, 1067–1075
42. Wang, K., McClure, J., and Tu, A. (1979) *Proc. Natl. Acad. Sci. U. S. A.* **76**, 3698–3702
43. Ohtsuki, I. (1974) *J. Biochem. (Tokyo)* **75**, 753–765
44. Wilson, F. J., Hirabayashi, T., and Cheng, M. (1977) *Tissue & Cell* **9**, 273–284
45. Ebashi, S., Iwao, O., Masaki, T., and Nonomura, Y. (1967) *J. Biochem. (Tokyo)* **61**, 817–819
46. Fowler, V. M., Sussmann, M. A., Miller, P. G., Flucher, B. F., and Daniels, M. P. (1993) *J. Cell Biol.* **120**, 411–420
47. Kruger, M., Wright, J., and Wang, K. (1991) *J. Cell Biol.* **115**, 97–107
48. Zot, A. S., and Potter, J. D. (1987) *Annu. Rev. Biophys. Biophys. Chem.* **16**, 535–559
49. Lehman, W., Craig, R., and Vibert, P. (1994) *Nature* **368**, 65–67
50. Sobue, K., and Sellers, J. R. (1991) *J. Biol. Chem.* **266**, 12115–12118
51. Vibert, P., Craig, R., and Lehman, W. (1993) *J. Cell Biol.* **123**, 313–321
52. Shirinsky, V. P., Biryukov, K. G., Hettasch, J. M., and Sellers, J. R. (1992) *J. Biol. Chem.* **267**, 15886–15892
53. Marston, S. B., and Redwood, C. S. (1991) *Biochem. J.* **279**, 1–16
54. Bogatcheva, N. V., Vorotnikov, A. V., Biryukov, K. G., and Shirinsky, V. P. (1993) *Biochem. J.* **290**, 437–442
55. Mak, A. S., Watson, M. H., Litwin, C. M. E., and Wang, J. H. (1991) *J. Biol. Chem.* **266**, 6678–6681
56. Mak, A. S., Carpenter, M., Smillie, L. B., and Wang, J. H. (1991) *J. Biol. Chem.* **266**, 19971–19975
57. Ikebe, M., and Reardon, S. (1990) *J. Biol. Chem.* **265**, 17607–17612
58. Moncman, C. L., and Wang, K. (1995) *Cell Motil. Cytoskeleton* **32**, 205–225
59. Labeit, S., and Kolmerer, B. (1995) *J. Mol. Biol.* **248**, 308–315
60. Wang, K., Knipfer, M., Huang, Q.-Q., Hsu, L., van Heerden, A., Browning, K., Quian, X., and Stedman, H. (1990) *J. Cell Biol.* **111**, 428a